NASA TM-80760

National Space Science Data Center

79-09

(NASA-TM-80760) THE SPACE SCIENCE DATA
SERVICE: A STUDY OF ITS EFFICIENCIES AND
COSTS (NASA) 56 p HC A04/MF A01 CSCL 05B

N80-19982

Unclas
G3/82 33602

THE SPACE SCIENCE DATA SERVICE:

A Study of its Efficiencies and Costs

December 1979

NSSDC 79-09

THE SPACE SCIENCE DATA SERVICE: A Study of Its Efficiencies and Costs

James I. Vette

National Space Science Data Center
Goddard Space Flight Center

Robert H. Hilberg
Donald E. Zuhl
Albert E. West

Sigma Data Services Corporation
Silver Spring, Maryland

December 1979

National Space Science Data Center (NSSDC)
National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt, Maryland 20771

CONTENTS

CONTENTS (Concluded)

1.0 Introduction

Within NASA's Office of Space Science (OSS) programs, a data management plan has evolved for increasing our knowledge of the natural processes affecting the world we live in. These programs are collecting the order of 10^{12} bits of data annually. These data, once acquired by the appropriate tracking network, are copied onto magnetic tape and sent to the appropriate principal investigators. After these scientists analyze the data for an appropriate number of years, some data are deposited in the National Space Science Data Center, which makes them available to other investigators for use in their research.

During the decades over which these procedures have evolved, the computing equipment available to investigators has changed radically. While equipment at the high end of the scale has become larger and faster, an even more important revolution has occurred at the low end. Small computers are much more readily available, to the extent that distributed processing and the use of intelligent terminals, both graphics and alphanumeric, have become extremely commonplace. Data transmission and communications are making techniques available that were impossible or prohibitively expensive 10 years ago.

NASA's needs have also evolved over the years. While OSS programs have been collecting approximately 10^{12} bits of data per year recently, missions are being planned that could increase this quantity significantly.

The concept of a centralized computing facility with a centralized data base accessed by a large number of remote users has been discussed, and a group to weigh the advantages and disadvantages of such a Space Science Data Service (SSDS) has been established. The task of this study group is given in the attached letter establishing the SSDC Study Group.

This report has been prepared to provide information helpful to the SSDS Study Group's consideration cf the best way to approach OSS's Data Management Plan. Four approaches, determined by the Study Group, have been considered: 1) the status quo, consisting of each of OSS's investigators using his home institution's computing facilities, and data being distributed by magnetic tape; 2) a centralized computing facility with a centralized data base that can be used by all OSS investigators at their home institutions using interactive terminals; 3) a centralized computing facility with a centralized data base that could be used by investigators at their home institutions using some distributed processing compute capabilities; and 4) two centralized computing facilities, each with its own data base, that can be used by investigators at their home institutions using interactive terminals. This last possibility could involve division of the overall data base into one collected by Spacecraft Tracking and Data Network (STDN) and the other by Deep Space Network (DSN).

In the current procedure, approach 1, individual research groups tend to work independently, thus incurring redundant costs. The Information Processing Division (IPD) distributes experimenter tapes, which were generated by doing preliminary processing on raw data records, to individual principal investigators (PIs) and sometimes to their affiliated groups (Co-PIs). The PIs acquire their computer resources independently, ordinarily a computer facility operated by their own institution to support its totality of projects. Some PIs, particularly Goddard-based

PIs, utilize the Science and Applications Computing Center (SACC). PIs are expected to submit analyzed data to the National Space Science Data Center (NSSDC for archival and availability to an interested scientific public.

There are three basic alternatives to the current mode of operation. Two alternatives that have been suggested would entail a centralized space science data base. Raw data would not be sent to the individual PIs, but would be retained at one or two centralized computer facilities. The possibility of two centers rather than one suggests itself because of inherent logistic reasons, namely the location of the two data collection networks (DSN and STDN) and their relation to their community of users. The alternatives as they have been proposed would offer both centralization of the data base and of the host computer facility. It is clear, however, that centralization of the data base would not require the centralization of the computer facility used by PIs. Thus, when we adduce the advantages and disadvantages of each alternative, we address these services separately.

A third alternative to the current mode of operation is to accord to PIs a distributed processing capability, where the FI could avail himself of a centralized data base and computer facility, yet maintain some local, probably specialized, computer resource.

A case can be made that suggests that cost savings will result from the centralization of the computer facility and the data base. Other factors, difficult to quantify, will also yield significant benefits to the local research groups when they have a large central computer facility at their disposal. These are the positive psychological factors that affect the individual researcher when he is working within a community of common interest. We are aware of possible negative factors, such as reluctance on the part of a PI who feels he might lose control over a local computer resource or over his data base. Since the NASA science program is one of mutual dependency between the Agency and the space science community it supports, the feelings of this community must be given considerable weight in assigning advantages and disadvantages.

2.0 Requirements

2.1 Data Analysis Requirements

In this section, two factors affecting the overall advantages and disadvantages of a centralized facility compared to decentralized computing will be discussed. These are the quantity of data that will be returned from space to OSS investigators, and the number and geographic distribution of these investigators.

Estimates of the data to be returned from OSS missions during the years 1979-1987 are given in Table 1 for the Ground STDN (GSTDN)/Tracking and Data Relay Satellite System (TDRSS), which is the name for STDN in the TDRSS era and Table 2 for DSN. A summary is given in Table 3. While these numbers reflect current expectations, some changes may be forthcoming for several reasons. Not all programs for the later years have been identified.

Recent experience has shown that data analysis costs average about \$1.83 x 10^{-5}/ bit. Further, about 400 floating point instructions are needed to analyze each bit with a hardware cost of about \$1 per instruction per second capability.

With the data returns projected for 1982, i.e., 2.6 x 10^{12} bits, the data analysis costs can be expected to be \$48 million. If all the planned, but not yet approved, missions are flown on their present schedules, the 6.1 x 10^{12} bits in 1986 would require about \$112 million to analyze the data to the level currently done.

The cost of copying data onto tape for distribution to OSS investigators is described in Section 6.2. It is shown there that the cost for distributing 10^{12} bits of data is approximately \$1.2 million, so that the cost of distributing 6 x 10^{12} bits becomes \$7.2 million.

Comparing these costs to the OSS expenditure for data analysis of \$55 million shows that some reevaluation must occur. While moderation in the amount of data acquired is likely, consideration must be given to ways of reducing data analysis costs.

2.2 OSS Investigators

Another key factor that will determine the requirements on a centralized system is the size and geographic distribution of the expected user community. For this purpose, the information in the tables is presented.

These data are based on the experiments and the responsible investigators listed in "Report on Active and Planned Spacecraft and Experiments, August 1978." These data include experiments that were active at some time during the period July 1, 1977 to June 30, 1978, or were planned as of June 30, 1978. While this selection is somewhat arbitrary, it is felt that it is representative of the body of experimenters that would be using the SSDS facilities over the next 5 years.

Table 4 gives the number of investigations and the number of groups performing these investigations broken down by type of measurement where possible. Included in this are some experiments no longer active (not listed in RAPSE) but for which active data analysis is being done. Table 5 lists the number of investigations broken down by PI affiliation and whether they are serviced by STDN or DSN. Those missions that will be handled by the Tracking and Data Relay Satellite System (TDRSS) are shown under STDN. These show that there are over 300 investigations carried out by over 100 groups at almost 100 institutions. Since GSFC investigations represent 20 percent of the total, one would expect the overall computer requirement to be approximately a factor of 5 greater than that of the SACC facility.

For a centralized computing facility, each of these institutions would have to be connected to the central facility through data communications links. The communications links to single users could be different from those to areas with a large number of users. From table 6 and figure 1, one can see that by linking four of the major cluster areas to a fifth, 75 percent of the users could be connected to the system, assuming that the SSDS would logically be located in an area where there are a large number of OSS investigators.

3.0 Levels of Support that May be Offered

We note three "levels" of support that SSDS may offer to its users. The level I user would use the computer facility through dumb terminals. At level II the computer facility would support intelligent terminals. Level III users would have a distributed processing capability in addition to access to the SSDS computer facility.

It should be stressed that in offering broader services to its scientific community, SSDS could provide different levels of support to individual users. Assuming that SSDS will support distributed processing, this is not a service that need be extended to all users. It has been suggested that in some sense, distributed processing is supported today; but no one who strongly advocates distributed processing as an advanced methodology would select NASA's user community as an example. We prefer to call the current environment decentralized rather than distributed. [SHE] Currently a user can be completely independent of a NASA computer center save for an interface that provides an experimenter data tape. The user frequently does not even observe the formality of returning reduced or analyzed data records to NSSDC. Access to a large central computer facility with a range of capability and equipment may currently be denied to the PI for various reasons. In fact, access to a large facility through an intelligent terminal may provide a greater computer capability than he now has.

The SACC facility now services level I users. SACC provides some support to NASA users who have a distributed processing capability through their laboratory minicomputers; however, this support is furnished to them as level I users. Laboratory minicomputers now serve functions that could be readily provided by an updated SACC facility. Each level of support would require both an updating of SACC computer resources and its data communications capability. The centralized facility proposed for SSDS would mean a fivefold increase in the customer base serviced by SACC today. The nature of the service offered by SSDS would also affect the data communications capability required. SSDS may support remote job entry and interactive processing through telephone lines or Value Added Networks (VAN), which are essentially "resale carriers" (DUN) of telephone data traffic. Users of SSDS may choose to transmit large volumes of experimental data. The telecommunications networks of the 1980s may provide less expensive alternatives to telephone carriers. Of course, NASA could implement through private industry a data communications network of its own. NASA could acquire leased lines, which would bring its PI-user group into the central facility that would constitute the SSDS. NASA uses this approach in trafficking data from its tracking and data acquisition stations to its centralized processing centers.

NASA intends to upgrade the current SACC facility. It is apparent that the planned capacity for FY82 prior to SSDS considerations would not handle the potential requirements of non-Goddard PIs, were they given the option to become SACC users. The Sciences Computation Needs Committee (SCNC) report estimates that an increased capacity of about .7 of a 360/91 would be required to handle the demand of 4,400 hours on SACC resources occasioned by the transfer of code 900 work and APL support to SACC after the release of the 360/95 by the Mission and Data Operations (MDO) Division. The SCNC report recommended that the SACC facility be upgraded to a capacity equivalent to three 360/91s. This was based on a total demand of 10,000 hours, a total capacity of 2.1 360/91s. This would mean that the upgraded facility would have an unused capacity equivalent to .9 360/91s or 5,656 hours. More recent planning has involved some aspects of an SSDS and the recognition that code 900 work will be handled by its own computer facilities in future years. Since no report is available for this, the SCNC report will be used.

If the demand for the non-Goddard-based PI is roughly equivalent to those of the Goddard-based PI, this unused capacity would not be sufficient to accommodate the additional demands on SACC resources entailed in the centralized computer facility envisaged for SSDS. This assertion is also based on presumptions in the SCNC report. The report estimates that the projected demand for code 600 users will be in excess of 4,000 hours. Since the Goddard-based PIs represent only 20 percent of all PIs, we assume that the non-Goddard users represent an additional demand of 16,000 hours on the computer resources offered by the SSDS. This, of course, exceeds that total demand of 10,000 hours that was projected for the SACC facility in FY82. The 16,000 hours represents the equivalent of 2.5 360/91s. Taking into account the .9 360/91 units unused capacity of the upgraded SACC facility as proposed in May 1978, the upgraded facility would still fall short of the potential demand by nearly 2 360/91s. Thus it is clear that, while the SSDS may be related to the SACC facility, since they would serve some of the same investigations, the SSDS is a new approach that must be evaluated independently of the more restrictive approach that is appropriate for upgrading the SACC facility.

3.1 Level I Support

The lowest level of customer could be serviced through dumb terminals. A DECwriter II or Anderson-Jacobson AJ 830 could provide a remote job entry capability and a degree of interactive processing on a central computer. Such terminals now service some SACC customers who can receive graphic output by directing output to the SD 4060 device. The 4060 facility will provide film and/or hardcopy output generally within 24 hours. Currently, some delay is occasioned by the physical separation of the mainframe and the 4060 facilities, and by the fact that the customer must intervene in order to transport the output of the SACC facility to the 4060 facility. SSDS could make this interface cleaner by causing 4060 input tapes to be automatically dispatched to the 4060 facility. The only potential time delay, then, would be in dispatching 4060 output to the Level I customer. Considering only terminal and modem costs for this level of support, the terminal could be purchased for $2,500 - 3,000 or leased for $120 - 130 a month with an additional $165 for purchase of the modem, or $10 monthly for its rental.

3.2 Level II Support

A second level of support could be provided for customers with intelligent terminals. Intelligent terminals would give the customers a remote job entry capability and permit, for example, interactive graphics. There is a broad selection of equipment that could provide the requisite intelligence. The Tektronix 4014 at a unit cost of $12,000 or the Tektronix 4025 at a unit cost of $7,000 are offered as a standard for comparison of costs and features. Most users would feel handicapped without a hardcopy printer. This would cost an additional $4,295 for the Tektronix terminals. The Hewlett-Packard 2647A Intelligent Graphics Terminal sells for $8,300. The terminal may interface with the HP9872A (Four-color Plotter) or the HP7245A (Thermal Plotter-Printer) through an "interface card" and cable.

3.3 Level III Support

There will be a category of user who will require a local compute capability above what is possible through intelligent terminals. Clearly, many of the current applications of minicomputers in the NASA data processing community could be transferred to a centralized computer facility. We could cite interactive graphics and program development as examples; however, this would ignore an important fact concerning this growth in minicomputer utilization. SACC did not provide an interactive graphics capability. Users saw the desirability of interactive graphics and acquired the resource that provided the capability. We note in our phase-in plan that SSDS should not expect to recapture this potential business until it establishes a "track record" for furnishing the support. This is also true in the case where minicomputers are used for program development. While probably no minicomputer was acquired expressly for this purpose, the minicomputer's performance no doubt encouraged broader applications. We are convinced that the programmers would not continue to develop programs on the minicomputer for later transfer to the host computer if it did not make their task easier. This use of the minicomputer will drop off in the environment of an SSDS facility.

A range of DEC minicomputer equipment may be used to illustrate possible costs. The DEC 11/34 has a basic cost of $10,310; it could easily cost over $100,000 when loaded with features. The DEC 11/70 is priced from $73,790 and the new DEC 11/780 is priced from $197,242.

4.0 Advantages and Disadvantages of Options

Four options for offering user support in the proposed SSDS will be examined. It was clear from the beginning that one aspect of two options was universally regarded as desirable; namely, the centralization of the data base. It is also clear that the centralization of the data base does not require centralization of the computer facility. It will be useful to address the advantages and disadvantages of these services separately.

The following are general observations concerning the advantages and disadvantages we have summarized: we cite a new cost entailed by SSDS as a disadvantage despite the fact that when all advantages and disadvantages are weighed, the new service may be cost beneficial; not all advantages and disadvantages can be accorded a cost; we do not contend that the lists of advantages and disadvantages are definitive, but feel that the lists may be useful in comparative analysis of the options. For example, some of the disadvantages of a centralized computing facility can be mitigated by a degree of distribution.

4.1 Advantages of Current Service

1. PIs may feel that they have more control over their computer resources than they would have over the computer resources of some centralized and geographically remote facility.

2. PIs have immediate access and complete control over dedicated decentralized computer resources. This may be desirable in the eyes of some investigators, but it has not been established whether or not this is to their advantage.

4.2 Disadvantages of Current Service

1. It is expensive and not cost beneficial to produce experimenter tapes and send them to PI sites. We have estimated that the annual cost is $1.2 million per 10^{12} bits of satellite data collected. The fact that copies of these tapes are made and stored by the processing facilities for two years is reflected in this figure. Using this figure, the estimated cost for CY79 is $3.4 million based on a total OSS collection of 2.8 x 10^{12} bits. If the data collected grows to 6 x 10^{12} bits for the future annual collection rate, the annual cost would then be $7.2 million.

2. Several experimenters have stated that a significant fraction of the total raw data they receive are never reduced and that much of the reduced data are never analyzed. This results from the lack of data analysis resources, although the desire to be able to do a much more complete job is high among the scientific groups.

3. There is a significant amount of duplication in both hardware and software. Preliminary estimates indicate that from 20 to 25 percent of the software and hardware utilization are duplicated. The hardware duplication occurs in both data storage and in the computer facilities utilized at the remote sites.

4. The user community's remoteness does not encourage the reuse of software code or exploitation of methodology by allied groups.

5. A change in an institution's computer center, where a NASA-funded investigator is a user, may require an extensive conversion or rewrite of software for which NASA incurs an additional expense for an essentially non-productive effort. Changes in the center's computer configuration or services are not controllable by either the institution-resident PI or NASA.

6. Reduced and/or analyzed data may never be deposited in NSSDC or the designated archive or may be deposited only after they have reached a low level of utility.

7. Some potential users could not use common facilities, even if they were available, without expensive revisions. Their individual systems may be designed around peculiarities of their local computer facility.

8. Experimenters tend to use the computer equipment that is avail-

able precisely because it is available, not becaus<span></span>e it is thu most efficient for their purposes.

4.3 Advantages of a Centralized Data Base

1. The development of a data base management system to handle the total OSS data base would be feasible.

2. The duplication, handling and mailing costs for experimenter tapes would be minimized.

3. Peer pressure is more likely to be applied to encourage more effort in reduction and analysis of data.

4. The trend towards what is termed synergistic studies, which utilize multi-experiment data, would be greatly facilitated.

5. NSSDC could automatically have access to data for distribution to secondary users after the proprietary period has elapsed.

6. Currently 16 groups involved in 66 percent of all experiments are concerned with data from both DSN and STDN. For these groups a centralized data base would be less expensive and more convenient than a bi-centralized data base.

4.4 Disadvantages of a Centralized Data Base

1. Some PIs are wary of maintaining proprietary data at a central location. If a secure data base manager was deemed necessary, it could be expensive. However, this fear seems unfounded since format and decoding information are generally necessary to access the data.

2. Centralization requires an enhanced data communications capability.

4.5 Advantages of Bicentralized Data Base

1. When compared to the current system, a system based on a bicentralized data base would have the advantages of a centralized data base listed in section 4.7, items 1 to 5.

4.6 Disadvantages of Bicentralized Data Base

1. Sixty-six percent of all investigations are now concerned with data from both DSN and STDN; consequently, these investigators would press for all their data in one place, or even worse, in both places.

2. PIs are wary of maintaining proprietary data at a central location. For those with data in both places, the problem is greater.

3. Centralization requires a data communications capability. Bicentralization would incur larger costs, at least for those users who need to access both DSN and STDN data.

4. It is possible that eventually the two facilities will not be able to communicate with one another because of differences in equipment configuration, operating system, etc.

4.7 Advantages of Centralized Computer Facility

1. This will give some PIs a powerful and efficient computer resource that may not be available to them now. A broader community of users would justify new services at the host, e.g., interactive graphics, color graphics, mass stores, COM.

2. This will preclude the problem of converting a NASA-developed software system were the non-NASA institution to change its computer configuration or services based on the overall good to all its users.

3. This will bring the non-NASA-Field-Center PI into a space-science community of users where he could share more directly the common experience of other PIs.

4. The users could share software and methods more readily. Although NASA maintains the GSFC Computer Program Library, as of June 1978 only 12 institutions have contributed to it. The reuse of software would save money, and it would tend to encourage further reuse when administrators can see that reuse yields results faster.

5. More efficient use of equipment should result. Capacity could be more easily shifted from one area to another, for example, when a mission or experiment fails. There should be less equipment idle time.

6. Some investigators are now using computer equipment to achieve their ends because it is the only equipment available, and it is not the most efficient for their purposes.

7. Compatibility among all users would be ensured.

8. Management of OSS computing expenditures would be improved by making them clearer and under a more centralized budget and control.

4.8 Disadvantages of Centralized Computer Facility

1. PIs may lose control over their minicomputer. Being one user in a large group of users will result in less individual control on the part of an investigator over his computing facility. While the user may see this as a disadvantage, it may produce no ill effects on his operation.

2. Centralization requires an enhanced data communications capability.

3. Security requirements may be demanded by some investigators.

4. SACC currently services approximately 1,000 users; centralization would increase the community to about 4,000. This fourfold increase in the user community would require management to take care not to understaff the facility with regard to support personnel and services.

5. Software conversion costs would be incurred by some PIs in moving to SSDS.

6. Individual groups distrust the ability of large computer systems to be responsive. This has been one of the reasons for the trend towards minicomputers seen at Goddard in recent years.

4.9 Advantages of Distributed Processing

1. A distributed resource, such as a minicomputer, may serve as a stand-alone facility or a remote job entry device to communicate with a host computer; thus the PI would have the flexibility to utilize his minicomputer for special purposes as well as have an interface to a powerful centralized computer facility.

2. Minicomputers over which the PIs have control and immediate access have psychological as well as processing benefits. Physical accessibility alone has a significant advantage beyond mere convenience.

4.10 Disadvantages of Distributed Processing

1. Minicomputers may be misused and are frequently under-utilized.

2. Control over activities at remote site is inherently limited.

3. Distribution requires an enhanced data communications capability.

4. Duplication in minicomputer resources increases costs of data analysis.

5. Utilization for common good or reconfiguration for new requirements is practically impossible.

5.0 Data Communications

The proposed SSDS will require an advanced data communications capability. There are several ways that NASA could meet the communications requirements entailed in linking investigators into the SSDS facility. NASA could contract with AT & T or with the VAN services to support SACC. Satellite communications holds great promise for the future; however, commercial tariffs have not yet been published, so a cost analysis is not possible. Facsimile transmission is discussed only to address the possibility that facsimile transmission might be used to support a network consisting of level I users of the SSDS facility. Packet-switching network services may be comparatively inexpensive for some applications, but under conditions of heavy traffic, leased lines are cheaper than public telephone lines, either through direct subscription or through the packet-switching networks.

The map (figure 1) offers a preliminary description of a communications network supported through leased lines. We have identified 11 major clusters of activity outside Goddard, viz., Los Angeles, San Francisco, San Diego, Boulder, Tucson, Ann Arbor, Houston, Boston, Hampton (LARC), Huntsville (MSFC), and New York. At each of these centers there are from 10 to 101 investigators. We have identified 43 other locations of lesser activity throughout the United States. Also shown on the map are NASA centers of excellence making up spread clusters, one consisting of the Universities of Minnesota, Wisconsin, Iowa and Chicago, and another consisting of the University of Texas at Austin and Dallas. In order to develop the SSDS network with optimally located concentrators, four important factors must be established:

1. The number of terminals at each location.

2. The amount of traffic between the location and the central site.

3. The cost of a low-speed line between each city-pair. Each cost figure would represent the potential cost of a line from one city to a concentrator in the second city.

4. The cost of a high-speed line between each city and the central computer facility. Each cost figure would represent the potential cost of a line from a concentrator or an isolated location to the central facility.

5.1 Packet-Switching Networks

AT & T plans to offer an Advanced Communications Service (ACS). ACS has been described as a "shared, switched data communications service." [RIN] The service would be "shared" in that customers would not have to maintain their own networks; they would configure "virtual networks" in the sense that computer time-sharing customers are offered "virtual machines." Initially, ACS would support only "general purpose terminals," viz., teleprinters, CRTs and remote batch terminals. Specifically excluded from the spectrum of supported equipment are the "special purpose terminals," e.g., plotters and facsimile transceivers. It is not clear what AT & T's long range plans are for supporting special purpose terminals. AT & T is expected to publish tariffs for service in mid-1979.

ACS has been proposed to meet the competitive data communications services offered by the VANs, the most pertinent one to our study being the packet-switching networks. Packet-switching networks represent a variation in and an advance over message-switching networks. The messages are packets of fixed length which are intended for rapid transmission over the network. The packet conventionally consists of 128 7- or 8-bit characters. Tariffs are expressed in 1000-packet sets or "kilopackets."

We interviewed a data communications expert who asserted that transmission rates in packet-switching networks are inherently limited (110 to 1200 baud); that, in fact, networks do not effectively service applications which require interactive processing. On the other hand, it can be argued that packet-switching is well-suited to an environment where processing is characterized by periodic bursts of activity, such as in interactive processing. In fact, the VANs currently do service customers with such requirements, e.g., the "large chemical company" that uses Telenet for "interactive computing." [LUR] The ARPANET is also used routinely for interactive processing. [KIM] [MAR2] The ARPANET is a direct predecessor of the Telenet service. It is interesting to note that ARPANET's design goal for its undetected error rate was less than one bit in 10^{12} bits [KIM], which is considerably better than the error rate for ordinary voice-grade circuits, and the total bits approximate our working number for the total collected by NASA annually.

We note two tariffs offered by competing packet-switching services. Datacomm has a $120 monthly service charge including equipment, and a charge of $1.70 per kilopacket during prime time (6 a.m. to 8 p.m.) and $.85 per kilopacket at other than prime time. Telenet charges an hourly rate of $3.25 for the "public telephone;" there is an additional charge of $.50 per kilopacket transmitted, and additional monthly charges for a microprocessor or modem and an X25 packet subroutine or other surcharged software. We estimate that a communications processor, supporting speeds up to 9,600 bps, and the X25 packet-switching interface would cost around $1,000 per month.

5.2 Facsimile Transmission

Some VANs specialize in facsimile transmission services, e.g., Graphic Sciences' Graphnet. Graphic Sciences also manufactures a line of facsimile equipment for non-network customers. General Dynamics' computer communications system is supported by the DEX 4100, which transmits a page in 2-3 minutes. This equates to 36 to 54 cents, or 18 cents a minute, over WATS lines. (STA3) The DEX 5100 represents an advancement over the DEX 4100. The DEX 5100 transmits a page in 20 to 90 seconds and rents for $320 per month, the approximate cost of the DEX 4100 in early 1978. The current price of the DEX 4100 is $179 per month. (ABR) Rapicom Corp. has demonstrated the feasibility of supporting an international teleconference with facsimile transmission of documents and translations. The Rapifax 100 is currently used on a daily basis by the United Nations. (STA2) The transmission time for a letter-sized document is 35 seconds. The company also offers the System 50, a microprocessor-controlled facsimile transceiver. It operates on voice-grade lines and transmits a page in 35, 50, or 90 seconds. The central-site unit rents for $475 a month; each remote unit rents for $340 a month. The central unit can be programmed to dial remote units, thus minimizing operator intervention. Facsimile has been suggested as the means of transmitting hardcopies of plots to level I users at remote sites. The monthly cost of $200 to $300 for a transceiver unit is small; but for an infrequent user it would represent a high cost per plot. Taking as a model a station with 10 users, each requiring 100 pages per month, these rates would produce approximate costs of $8,630 using DEX 4100, and $7,100 using DEX 5100 or Rapicom System 50. The DEX 4100 system becomes less expensive for loads of less than 52 copies per month, assuming 10 users per installation.

The 1980s are likely to see fairly dramatic advances in facsimile technology. In fact, the advances to date have been surprisingly slow. Facsimile transmission was used by the military in the late 1950s, but did not become widely used in business until over 10 years later. Today businesses are willing to accept transmission speeds of 2-6 minutes per page, but Satellite Business Systems (SBS) expects to offer facsimile transmission speeds which are 20 times faster than today's speeds. [KIN] The SBS service will not become operational until 1980 when its first communications satellites will be launched. The company is not in the position to publish a tentative tariff, but we assume that the competition among the communications-satellite companies for a potentially large user group will make their services "competitive" with those of the VAN services.

5.3 TV Broad-Band Transmission

Several vendors are offering or intend to offer a transmission facility over the 12 to 14 gHz bands, which have been allocated for both data and TV transmission. Satellite Business Systems (SBS) is scheduled to provide the service in 1980; however, the tariff has not been published. AT & T Long Lines, which may be providing a TV transmission facility to NASA today through Comstar, intends to replace Comstar in 1983. (WHI) Presumably, the plans call for broadening the company's data and TV transmission service by furnishing 4 to 6, 12 to 14, and 30 to 40 gHz transponders for major cities, the latter band being allocated for TV alone.

TV transmission or fixed-image transmission via TV bands as an alternative to facsimile transmission does not appear to be viable today for most potential commercial users. Forecasts for this service have not been reliable, but the networks which would offer the service appear to be optimistic that the service will create the demand. NASA is ideally positioned to take advantage of the technology when it becomes commercially viable. We understand that in 1977 Western Union contracted with NASA to furnish its Advanced Westar service. The intent was "to provide tracking and data relay service for all of NASA's earth-orbiting satellites." Western Union provides both 4 to 6 gHz and 12 to 14 gHz band service to its commercial customers through Westar. (WHI) TV transmission may prove to be a low-cost and reliable alternative for data transmission. These points have been advanced in behalf of PROMIS, a health information system funded by DHEW (WAN), but they have not been demonstrated. PROMIS has achieved satisfactory results over a 20-mile network. Experiments with fiber optic links have also had varying degrees of success over short distances. There is no doubt that there will be important technological breakthroughs in the 1980's, but it is too early to predict their costs and benefits to SSDS. We note that Western Union is currently under contract to study the feasibility of transmitting data in the 18 and 30 gHz bands by way of satellite during the 1980-2000 period. It is apparent that SSDS may profit from the eventual technological breakthroughs.

SBS is expecting to offer data transmission rates of from 2.4 kbs to 6.3 Mbs. The higher rate is 656.25 times faster than the 9.6 kbs which is a conventional line speed today. For purposes of comparison, note the calculation of 10.22 hours for the transmission of one fully-packed data tape at 9,600 bps. (CF. Section 6.2.) The same tape can be transmitted in 56 seconds or .015 hours at 6.3 Mbs.

6.0 Selected Issues Examined

During our study we were asked to address certain issues that were significant in the study of the proposed SSDS. We have attempted to find a basis for determining potential savings in software and storage. We addressed the issue of security precautions, which we feel will be of some consequence to PIs. We also offer a phase-in plan for users of the SSDS.

6.1 Software Savings

System Development Corporation (SDC) has performed an extensive study of software costs. The study acknowledged a large variance (58 percent) in the hours required to complete programs, and acknowledged that many subjective, unquantifiable factors apply; however, the study has served as the basis for subsequent attempts to develop a tool for estimating software tasks. It should be recognized that developing a reliable tool is a non-trivial task. GTE Data Services performed such a study over a period of 6 years without developing a unit of measure for programmer productivity. The findings of the GTE study were prefaced with the following caveat: "It is certain that we have not completed the process of data processing measurement even though we have been at it for 6 years." (PEE)

For our calculations we will use SDC's mean figures of 5.89 manmonths for 1,000 lines of "machine-language" code, and 2.13 manmonths for 1,000 lines of "procedure-oriented-language" code. (SHA) Assembler Language (ALC) falls within the definition of the former, and FORTRAN within the definition of the latter.

In the course of interviewing potential SSDS users, we encountered the argument that the savings in software that would be realized in SSDS would be insignificant in comparison to hardware and operations costs. Two scenarios from a university computer center were depicted. In the first scenario, original software was developed for the data reduction and analysis under a NASA-sponsored project. In the second, the data reduction and analysis programs were rewritten when the university's computer center changed equipment. In both cases, it was said, the software development costs were insignificant when compared to hardware-related costs for the system in production.

The typical data processing budget shows a significantly higher percentage of costs planned for personnel than for hardware capital investment. One recent survey suggests that the typical breakdown would be 53 to 32 percent. (DOR) The difference is likely to increase in time, since wages are increasing while we are currently experiencing some reductions in hardware costs. Coincidentially, declining hardware cost is one of the conventional reasons adduced for the "distributed trend." The current rate of decline has been said to be 30 percent per year. (STA1)

It should be noted that it is not a unanimous opinion that the relative cost for personnel will increase. It has been argued that salaries currently account for 35 percent of DP budgets and they will account for only 26.5 percent by 1985. (STA5) The argument is based on the assumption that the use of software and firmware will reduce the need for people. It is also assumed that users will be caught up in the candy-store syndrome, where they will be unable to resist the selection of available equipment.

If we use the SDC line estimates with some of the figures elicited from the questionnaire, we can make some gross calculations for potential savings in the cost of software that would accrue from a centralized computer facility where software is freely shared. We considered 6,000 lines of code for the HEAO X-ray astronomy experiment. It is estimated that 5 percent of the lines were written in ALC; thus there were 5,700 lines of Fortran and 300 lines of ALC. Using 2.13 manmonths per 1,000 lines of Fortran code, we calculate 1942.4 programmer hours. A survey of GSFC investigators (see section 6) showed that about 20 percent of higher level code could be used by other groups. Assuming this 20 percent of reusable code to be valid, we have a potential saving of 388 hours. Taking an hourly rate of $20 for a programmer/analyst, the saving would be $7,760 out of a total expenditure of $38,800. We found that 100 percent of the coding in ALC tends to be reused; thus at 5.89 manmonths per 1,000 lines of code, the comparable time for 300 lines is 283.2 hours or $566. Therefore a savings of $8,332 would be available after an expenditure of $39,366, or 21.2 percent.

We reviewed the data reduction and analysis system for an IMP magnetometer experiment. Of the total of 6,955 lines of code, 3,445 were in Fortran and 3,510 in ALC. We calculate that 1,174 hours may be expended in implementing 3,445 lines of Fortran code. Assuming again that 20 percent of this code is reusable, the potential saving is 234 hours or $4,680 out of a total of $23,400. Assuming that 100 percent of the 3,510 lines written in ALC is reusable - which it was within the system - then the saving would be 3,339 hours or $66,780. Thus in this investigation, expenditures of $90,180 were made that would have produced reduced expenditures of $71,460, or 79.2 percent for other projects.

6.2 Storage Savings

Experience at NSSDC over the last 10 years has shown that the following factors go into the cost of reproducing a tape and sending it to an investigator. Computer time depends on the machine used, but estimates range from $11 for the NSSDC MODCOMP IV computer to $14.40 for the SACC 360/75, and $28.75 for the SACC 360/91, so that $14.00 can be taken as a fair estimate. Labor, including cataloging, setting up and submitting computer runs, handling the tapes, and documentation through shipping and storage, averages about $13.00 per tape. The tapes themselves cost $10.00 each, and mailing costs $3.00 per tape. This gives a total cost of $40.00 per tape.

About 4.4 x 10^7 bytes may be stored on one 1,600 bpi tape (SHA), and 2.2 x 10^7 on one 800 bpi tape. In practice, IPD sends out tapes that contain about 40 percent of this value, because of inter-record gaps and unfilled reels, and the average density is 1,200 bpi because some users still require 800 and 556 bpi. We estimate that around 15,000 tapes are required to store 10^{12} bits of data acquired annually. Because of orbit/attitude data and the addition of some other experiments on an experimenter's tape, there is a data volume amplification of about a factor of 1.5 over that received at the data acquisition station. Since two copies are made, one for the investigator and one for retention at GSFC, this total becomes 30,000 tapes.

These totals produce a total cost of $1.2 million. IPD should be able to provide more precise figures. Based on the expected data return shown in table 3, this represents an annual cost of $3.4 million in 1979 rising to $7.3 million in 1986. These amounts are a significant fraction of the total expenditure for data analysis.

SSDS may be pressed to offer more random access memory (RAM) to its user community; however, we found that it was difficult to estimate demand based upon our survey of users. Most Goddard-based users of SACC appear to have developed systems which use tape almost exclusively. When disk space was used it was used for temporary storage of less than one cylinder. One user suggested that 1,000 cylinders would be "useful." One user did suggest that around 10 percent of his data was disk-resident, but there was no estimate of the size of the data base. The system involved is recognized as an exception at Goddard. Clearly it is difficult for users to speculate on storage requirements, given that more mass storage is available. Projecting a demand will have to await further analysis, but we offer the figure of 1 percent of the total collected by NASA yearly, or about 10^{10} bits of storage.

The cost of storage has dropped significantly in the last few years. The 4k dynamic metal oxide semiconductor (MOS) chip is the "most widely used memory component today, with typical access time ranging from 100 to 500 nanoseconds at the cost of about 0.1 cents per bit." (STA7) Auerbach expects RAM technology to provide a 64k chip for about 0.05 cents per bit in the 1980s. This compares favorably to the 2.5 cents per bit of 1k RAM chips in the early 1970s. The figure of 0.1 cents per bit appears to be a good working figure for IBM 303X or plug-compatible storage. (STA8) Using this per-bit cost, storing 1 percent of the total bits collected by NASA yearly would cost $10 million worth of 303X storage. An annual maintenance cost of 5 to 15 percent has been estimated. (SHA) (YAS) The savings in tape duplication, distributing, and storage would be partially offset by such RAM storage requirements.

If SSDS utilized its data communications network to transmit significant amounts of data, other costs would be incurred. Consider the time and cost for transmitting one 2,300-foot 1,600-bpi tape at 9,600 bits per second:

9,600 bits/sec = 1,200 bytes/sec
2,300 feet x 12 inch/feet = 27,600 inches/tape
27,600 inches/tape x 1,600 bytes/inch = 44,160,000 bytes/tape
44,160,000 bytes ÷ 1,200 bytes/sec = 36,800 bytes/sec or 613.33 min. or
10.22 hours.

The hourly charge for a selected packet-switching network is $3.25; thus the charge for service less the per-packet charge is $33.22. Since there are 128 characters in a packet, 44,160,000 bytes is 345,000 packets or 345 kilopackets. The per-kilopacket charge for the selected network is $0.50; thus the full charge in the example is $172.50. Therefore, the total charge, $172.50 + $33.22, would be $205.72.

Assuming that NASA leases a line, e.g., between Goddard and Chicago, and that the line is fully used to transmit data so that a realistic cost basis can be established from a unit of time in which the line will be utilized, then we can make a comparison to the packet-switching cost. The lease cost for the Goddard-Chicago line would be $1,000 per month. Considering then the time distributed over one full shift of 8 hours:

$1000/160 hours = $6.25/hour

The cost for 10.22 hours would then be $63.97. Considering two full shifts, the hourly rate would be $4.16 and the cost for 10.22 hours would be $42.51. The corresponding figures for three shifts are $2.08 per hour and $21.25 for 10.22 hours.

The cost for transmitting the 10^{12} bits of data through the packet-switching network would be nearly $6 million. Using leased lines, the cost would be somewhat over $600,000; roughly equivalent to what we have estimated the present cost for duplication and the purchase and mailing of tapes. It is obvious that it is less expensive to transmit a large volume of data continuously over leased lines than to transmit the data through a packet-switching network. Packet-switching is more economical when processing requirements can be accomplished in bursts of activity, specifically when the user's total communications costs are less than the cost of the leased line. These costs are given only to set a scale for this type of data transmission. One would expect relatively little data to be distributed to SSDS investigators since the basic idea of an SSDS is to provide centralized processing colocated with the central data base.

6.3 Security Precautions

With any degree of centralization in the services offered by SSDS, the PIs will voice a concern over system access security. In the current procedure, IPD sends raw data tapes to PIs. Since the PIs physically hold these tapes, they have felt that they can exercise a degree of control that is unavailable to them when the raw data tapes are retained at some central facility. SSDS must maintain adequate security procedures to satisfy the PIs.

Access to the individual data files could be password- and account number-protected. However, experience at NSSDC has shown that unless documentation is available, meaningful use of an investigator's data by an outside person is extremely difficult, to the point of being impractical. The access to data can be controlled very effectively through the release of documentation.

6.4 Phase-In Plan

The phase-in plan for providing SSDS support will be controlled by the equipment procurement schedule and by the support it receives from the user community. Because of the significant digression from existing operational techniques, a prototype system to gain some experience would be helpful.

Such a trial period should serve approximately 25 investigators using, perhaps, a super minicomputer, and should last one year. The experience gained during this trial would be extremely valuable in designing the overall SSDS system. It would also verify the validity of the centralized system concept and build confidence within the user community.

After this trial period, a realistic phase-in plan could be devised. However, other information regarding the requirements of potential users must be researched during this period. This information must include the following types of facts:

1. NASA may require all new investigators to utilize the services offered by SSDS. How many such investigators, and their locations and projected workloads, must be identified.

2. Level I users, and level II and level III users whose software is transportable to the SSDS computer facility, could be supported immediately. Their requirements must be identified.

3. Users whose software requires some conversion before it is compatible with the software offered by SSDS would be supportable by SSDS according to an indeterminate schedule. SSDS can support users as their software becomes transportable.

4. Users who require special processing considerations, such as interactive graphics, will require supporting software at SSDS. The success of the SSDS depends on the support of its users so that SSDS must establish a "track record" to persuade other users to forsake their existing systems.

5. Users who have a substantial investment in non-compatible software at their current computer facility may find it difficult or impossible to transfer processing to SSDS. The cost to convert must be compared to the benefits to the user in the SSDS' services. This is further complicated by software systems which were designed around peculiar hardware. Such users may never convert to the SSDS system. Only after details such as these regarding numbers, locations, workloads and timescales of conversion for the overall group of potential users are known, can a realistic phase-in plan be devised.

7.0 Questionnaire Analysis

The SSDS Study Group suggested that a questionnaire be devised that would elicit information on the current mode of processing and on the potential impact of changes resulting from conversion to an SSDS. The questionnaire provided in

Exhibit A was derived and distributed to a number of GSFC user groups, including codes 626, 664 and 694.

The following are general observations on the individual questions:

Q1 · Only one response was recorded, namely 168 hours.

Q2 - Only one response was recorded, namely an estimate of 936 hours per instrument per year.

Q3 - This question did not encourage unbiased answers.

Q4 - We found that 20 percent is a useful figure for estimating a percentage of common software. Our confidence in this figure increases if we are to include software that is readily adaptable to other groups and software techniques that have broad applicability.

Q5 - The interviewees were not prepared to estimate the number of groups which could conceivably utilize the software they had developed. Thus, it is one thing to say, as for Q4, that 20 percent of the software should be reusable, but it is quite another to say who will use it.

Q6 - Those interviewed seem to accept the figure of 80 percent as a good working figure for sensor data. It was the independent estimate of several of the interviewees and subsequent investigation increased our confidence in the figure.

Q7 - The question reads "By how many other groups is the typical bit of your non-sensor-output stored data held?" This has been interpreted to mean, "How many other groups use the non-sensor data contained on your raw data tape?" The answer to this question did not seem to be generally known.

Q8 - The question elicited an estimate of software in comparison to data storage. It was universally felt that the proportion of storage accorded to software was so small as to be insignificant. The question also asked for a percentage of the software used by a group that was unique to the group. Because of the way the several systems have evolved, 100 percent would be a legitimate response; however, most of those interviewed accepted a working figure of 80 percent.

Q9 - Responses to this question vary greatly. One system uses the Sigma 9 exclusively, one uses the SACC 360s exclusively. One system uses the 360s into the data analysis phase before additional analysis is performed on laboratory minicomputers. The only percentage-breakdown that was offered estimated that two-thirds of the work was performed on SACC computers and one-third on laboratory minis.

Q10 - It is not customary to generate job accounting statistics for a computer-utilization report for minicomputers. One programmer suggested the minicomputer runs 90 percent of the time, implying virtually a full shift. A programming manager indicated 4, 5, and 11-1/2 shifts per week on three different minicomputers.

A1 - The responding investigators reported that their data tapes consisted of approximately 80 percent of their own data. The percentage accorded to data flags, time, housekeeping, and command data was insignificant in comparison. The time data would presumably be the basis upon which a thread would be strung between schema. It would be possible to maintain attitude/orbital data as a discrete schema where the data would be accessed through the separate sensor-data schema. This appears to have been the design goal for the AE system, which was surveyed.

A2 - The SACC 360 users we encountered use tape files, using disk for only temporary storage. It seem clear that the reason for this is that SACC does not make it convenient for users to maintain disk files. When permanent disk files are maintained, they are used at the PI's minicomputer.

A3 - Most PI's do not duplicate the IPD experimenter tape, but will retain one copy of reduced and analyzed data. The one exception to this was the X-ray astronomy group which maintains two copies. This is appropriate since programmers can alternately use the laboratory minicomputer or the SACC facility.

A4 - The analyzed data tapes are generally kept from 3 to 5 years. We found no one that retained a copy of the experimenter tape.

B1 - Responses to the B-series of questions were not instructive. The responses to B1 told us little more than that data reduction and analysis programs reduce and analyze data.

B2 - Responses suggested that all data were used, even from one area where it was suggested informally that the raw data tape carries a large portion of extraneous data.

B3 - We could not perceive a pattern in the reduction or expansion of data.

B4 - The typical program will produce a restructured data tape for subsequent processing within the system and one or two tapes to be input either to a film processor or plotter/printer.

C1 - We found Fortran to be the only higher-level language used.

C2 - A broad spectrum of equipment was used: the SACC 360s, the Sigma 9, and several minicomputers, including PDP 11/70, PDP 11/34, IBM 1800, and Interdata.

C3 - The 360 programs tended to be large, typically 200 to 300k, with one reaching 500k. One plausible suggestion for this is that "core is free," as one programmer said. A job will not be placed into a restricted class until it exceeds 500k on the 360/91 or exceeds 300k on the 360/75. We found no large disk usage. We believe this is because disk space is expensive and apparently requires adherence to administrative procedures which programmers would prefer to avoid.

C4 - Programs use from one to five tapes, temporary disk storage, and the systems printer. We found one system which uses the card reader and punch. The number of tape drives is not in itself informative. A frequent use of a tape output is to provide input to a microfilm printer or plotter/printer.

C5 - We found that interactive graphics were widely used and probably would be in wider use if they were more readily available. We are not prepared to say that any user "requires" an interactive processing capability; it is perhaps in the category of a service that creates its own demand. We would argue, however, that for a group to start to use this capability would cause more than a software conversion; it would require a change in overall approach.

C6 - In only one system did we find extensive use of library functions; however, this was the only system where we examined the program listings ourselves. We suspect that many programmers use library functions so routinely that they may not distinguish between the functions and Fortran statements.

C7 - FTIO, the Fortran I/O package and the SD4060 interface package appear to be the most widely used packages. Other unspecified packages were used for bit manipulation and matrix inversion.

C8 - A typical program consists of 2,000 lines of code, 15 to 50 percent of which may be preexisting.

C9 - It is very difficult to generalize on the percentage of a program which may consist of assembly language code. We received the following responses: 5, 10, 15-20, 20, 34, 55 and 60. We believe it to be true that assembly language code tends to be reused.

C10 - As we noted above, routines written in assembly language tend to be reused. They tend to be written for a special function or purpose, e.g., to perform time series analysis, curve fitting, translate between coordinate systems, statistical analysis, to generate data bases and numerical tabulations.

C11 - No one who was surveyed was involved with a system where programs are compiled in production runs.

7.1 Questionnaire - Magnetometer

The magnetometer system reduces and analyzes data collected through IMP-J. No answers were provided for Q1 and Q2; however, the group would probably argue strongly that very little work could be shifted from the laboratory's minicomputer to the SACC facility (Q3). One of the primary reasons for this is that the minicomputers serve special purposes, one of the most important being that they serve as a tool for instrument design.

We have had occasion to discuss the commonality question with another group using IMP-J data, and we are prepared to accept the judgment that the two groups would not have been able to share software (Q4). We are bound to say, however, that the lack of commonality may be due to program structure; thus, if you are pre-determined to reuse code, some portion can be reused. In the magnetometer programs, for example, some assembly language coding for plot generation was reused. Presumably it could be reused outside the group as are generalized plot packages (Q5).

We found that 80 percent is a good estimate for the percentage of sensor data in the system's data files (Q6); the bits by type that may be held by other groups are the orbital, attitude, and other data consisting of approximately 20 percent of the experimenter tape (Q7).

In reviewing the system's file layouts, we made the precipitous judgment that the percentage of sensor data seemed to decrease as the data structure was revised in successive steps of reduction and analysis. In fact, an analyzed data tape did have a significantly smaller percentage of sensor data and a significantly higher percentage of data derived from orbital data. Since these non-sensor data could be derived again, we could argue that they need not be retained with the analyzed magnetometer data. We have concluded, however, that the cost to derive this data each time they may be required would be greater than the cost to store it with the sensor data.

The storage requirement for software when compared to data is insignificant, and we have indicated before that we are not prepared to say what percentage of the magnetometer software is not unique to the group (Q8). No answers were provided for Q9 and Q10.

The following is a bit-breakdown by data type for the experimenter tape, a "detailed" and a "summary" data tape, coresponding to a reduced and analyzed data tape (QA1):

	Instrument	Data Flags	Time	House-Keeping	Attitude/ Orbital	Other
Experimenter	74.85	0.3	0.7	1.0	8.6	14.5
Reduced	92.6	0.7	2.2	3.6	0.4	0.4
Analyzed	32.3	1.5	4.4	4.4	25.0	32.4

The summary data tape contained about one-tenth the number of bits of the experimenter tape, while the detailed data tape contained about 4 times as many bits as the experimenter tape.

The system uses tape as the primary storage medium (QA2); disk is used only for temporary storage at SACC. No tape backup is made of the raw or reduced data tape, but backups are taken of the system's "detail" and summary" tapes (QA3). The summary tape may be retained for only 1 month, however, and the detail tape for a least 5 years (QA4).

Programs convert "experiment data to field vectors in physical units," transform data from spacecraft coordinates to solar ecliptic and solar magnetospheric coordinates, and from a 0.040-second time frame to a 1.28-second time frame (QB1).

The magnetometer programs typically use flag, attitude, time and housekeeping data as well as experimental data (QB2). It seems the command data are not viewed as an entity apart from sensor data.

There is no general observation appropriate for the reduction or expansion factor. We noted programs that expanded data by factors of 1.1 and 4 and one program that reduced data by a factor of 41 (QB3). The typical program produces tapes for microfiche (for printouts), microfilm (for plots) and for the restructured data or statistical summaries. One program generates "hourly average cards," which evidently are reintroduced as input within the system (QB4).

Fortran is the higher-level language used in the system (QC1), but assembly language code may take up from 10 to 60 percent of a program (QC9). Assembly language code was used for plotting routines and other special purpose, frequently used operations (QC10). The programs that we examined run on the SACC facility's 360s under OS/MVT (QC2) and require from 200 to 330k of core storage (QC3), up to four tape drives (QC4), one track of temporary disk space, and the system printer (QC4).

Interactive processing is not performed (QC5); we assume this is because the SACC facility does not support this capability. We note that programs use from 7 to 14 library functions, e.g., COS, SIN, SQRT, etc. (QC6) and two generalized service packages, viz., the SD4060 plot and FTIO, Fortran I/O package. Up to 50 percent of a programs's code may be pre-existing (QC8); this is largely due to the fact that the group has developed some reusable or easily adaptable code for plot generation. No programs are compiled in production runs (QC11).

7.2 Questionnaire - Mass Spectrometer

The system was developed for data collected by the Atmosphere Explorer (AE) series of spacecraft. The system is characterized by two features that make it unique among the systems reviewed: it was wholly resident on the Sigma 9, and the system had some of the features of a data base management system (DBMS). Because of this latter design characteristic, the answer to question QA1 is stated in a detail that was not feasible for any other group. The following is given as the logical structure for the AE raw data tape:

	Total	Experimental Data	Data Flags	Time	Housekeeping	Command	Attitude	Orbital
	100	73	1.0	2.0	2.0	2.0	10	10
Own Experiment on S/C	10	7.3	0.1	0.2	0.2	0.2	1.0	1.0

	Total	Experimental Data	Data Flags	Time	Housekeeping	Command	Attitude	Orbital
Other Experiments of Interest on S/C	60	43.8	0.6	1.2	1.2	1.2	6.0	6.0
Other Experiments of No Interest on S/C	25	18.25	0.25	0.5	0.5	0.5	2.5	2.5
Correlativ Data - GBR/ other S/C	5	3.65	0.05	0.1	0.1	0.1	0.5	0.5

The group's own sensor data is then viewed after two steps of processing:

	Experimental	Flag	Time	Housekeeping	Commands	Attitude	Orbital	Other
Step 2	29	6	12	11	1	19	7	15
Step 3	90	3	3	0	0	0	2	2

The nature of the "other" data was not specified, and while it represents an insignificant percentage of the analyzed data, it was a comparatively large percentage of the intermediate file.

AE data are stored on both disks and tapes. About 10 percent of the raw data is stored on disk. The analyzed data are maintained on disk (QA2). One tape backup is made for disk files (QA3). Raw data are retained for 3 years, and the reduced and analyzed data for 5 years (QA4).

The programs calibrate and adjust data depending on attitude and orbit. The programs list, plot, and reformat experimental data into machine-sensible data. Some non-sensor data are listed, plotted, and stored in the machine-sensible data structure (QB1). Virtually all data types are used by the programs: attitude, orbital, time, housekeeping, commands and data flags. These data are used to modify the sensor data (QB2). Obviously something is done with the "other" data, but this is not indicated. Three reduction factors are given for data of three types: Sensor 7:1, O/A 20:1, and Time 1:1 (QB3). Programs generate 13 pages of printout and 17 pages of plot output (QB4).

Fortran is the higher-level language used by programmers (QC1). The system uses the Sigma 9 with the CP-V operating system (QC2). Programs have a core requirement of 130k and a disk space requirement of 8 million bytes (QC3). Programs use one tape drive, a printer and a disk (QC4). Since a run requires three tape mounts, the Sigma 9 configuration may be viewed as a constraint on the system. The system averages 5 runs per day, thus 15 tape mounts per day.

There is no interactive processing in the system (QC5). Plots are produced through the offline SD4060.

Some unspecified number of Fortran scientific functions are used (QC6) and no generalized service packages are said to be used (QC7), although we assume that "Wolfplot" is used in connection with the system's SD4060 interface. The system consists of 6,400 lines of code, none of which were pre-existing (QC8), and 2 percent of which are in assembly language (QC9). Assembly language code was used for routines that perform curve fitting. Some ALC routines are said to "interpolate" or "calculate" (QC10). No program is compiled in production runs (QC11).

7.3 Questionnaire - X-Ray Astronomy

The X-ray astronomy system was developed for HEAO 1 data. The system extracts data from 180 6,250 bpi tapes. After reduction and analysis, the experimenter tapes are returned to IPD. No tape copies are made since much of the data on the experimenter tape is considered extraneous.

No direct answer was recorded for Q1-Q3. In fact, work is logically distributed between the SACC computers and the DEC 11/70 in the laboratory. The DEC equipment is often used for program development, the resulting programs being subsequently transferred to the SACC facility. The programmers' point of view of Q3 is, "What can be shifted to the SACC facility is shifted as a matter of course." It would be argued that program development as a function could not be shifted, nor could the graphics capability the laboratory now has. In fact, the transportability issue may be further complicated by the structure of the programs since they have evolved due to the nature of the available equipment. We expressed surprise over the size of the 360 programs. (We also understood the total size of the system to be 19,000 lines of code.) It was suggested that the 360 programs tend to be large because "core is free" and the CPU time is not. Programs run on DEC equipment, on the other hand, tend to be small since core is limited.

It was estimated that about 5 percent of the software developed (Q4) for the system could be used by other groups. The reusable software is in the form of common subroutines and represents an investment of approximately 20 manhours per year. Approximately 10 scientific subroutines (Q5) could probably be used by the group investigating related balloon-flight experiments. Common subroutines, scientific and mathematical, and some special packages, especially for plotting, are stored and readily available to all programmers in the group (Q8). Software probably takes less than 1 percent of the storage space available to the group (Q8).

It was estimated (Q6) that 80 percent of 10^{10} bits of data maintained by the group results directly from the group's own sensor data. Apparently three other groups were interested in the non-sensor data contained on the experimenter tapes (Q7).

Approximately two-thirds of the group's software runs on the SACC computers and one-third on the laboratory's minicomputer (Q9). Approximately 10 percent of the software may run on both, but in fact the mini is used for development of software to be run at SACC. The mini is also used for special processing that

SACC does not adequately support. The laboratory's minis run approximately 90 percent of the time when programmers are working (Q10), or approximately five 8-hour shifts per week.

The following bit-breakdown was offered for raw data tapes (QA1):

Experimental Data	50%
Data Flags	10%
Time	0.1%
Housekeeping	5%
Command	5%
Attitude	20%
Orbital	5%
Other (e.g., environmental)	5%

Tape is the basic storage medium (QA2); however, some disk files of analyzed data are retained, the largest of them being six tracks in length. No tape backup is taken of the tape in IPD format, but two backups are taken of the group's reduced and analyzed tapes (QA3). Two copies are taken because the group programmers may use both the SACC facility and their own laboratory minicomputer. The group's reduced data tape is retained for 5 years.

QB1 and QB4 elicited no useful information. Since the group is not satisfied with the IPD format, as a part of data reduction, the data are selected out and reformatted. One pass of the entire data file of 180 tapes is made just to examine the data. An analytic report and plot are produced - the group makes extensive use of the SD4060 and, in fact, was responsible for improving the Wolfplot software since they were dissatisfied with the efficiency of the package.

The system's programs use attitude, orbital, flags, housekeeping, and command data as well as sensor data (QB2). The sensor data are expanded roughly 50 percent; no other data type is expanded (QB3). Some housekeeping data are dropped, and some data selection flags are added, e.g., those reflecting earth occultation, status and bit error data.

Fortran is the higher-level language used by programmers (QC1). The DEC 11/70 uses the RSX-11D operating system (QC2) and is well-loaded with 2 tape drives, 224k words of core, 1 disk with 88 Megabytes of storage, a graphics device with a hardcopy printer, and a line printer. Programs run on the 360 take from 150 to 550k; disk space of up to 1,000 cylinders or 180 Megabytes would be "useful" (QC3). We note that the core storage requirements entail some restrictions on running at the SACC facility. Programs need up to four tape drives for some purposes, and the system has been developed with a degree of online processing (QC4), especially for interactive graphics (QC5). Programs use two library packages: FTIO (Fortran I/O Package) and the Wolf Plotting and Contouring Package (QC6) and some generalized service packages for matrix inversion (QC7). Programs have up to 6,000 lines of code with approximately 40 percent of the code being characterized as "pre-existing" (QC8). Around 5 percent of a typical program is written in assembly language (QC9). Assembly language code is used for routines that do curve fitting, time series and least squares analysis (QC10). No programs are compiled in a production run (QC11).

7.4 Questionnaire - Energetic Particles

It was estimated that 168 hours of SACC computer time were used for each instrument per year (Q1). The estimate for minicomputer utilization was 936 hours per year per instrument (Q2). A breakdown was provided by minicomputer:

DEC 11/70	92 hours/week for 5 instruments
DEC 11/34	40 hours/week for 0 instruments
Interdata	32 hours/week for 3 instruments

Very little that is now done on the laboratory's minicomputers can be transferred to the SACC facility. One of the reasons for this is that the laboratory makes extensive use of the interactive graphics capability of their minicomputers; however. it was estimated that up to 2 percent of the software could be transferred (Q3). Less than 10 percent of the software developed for the project surveyed could be used by other groups. No estimate could be provided for what this would equate to in manhours (Q4). No estimate was offered for the percentage of software lines unique to the group; but the storage required for software when compared to data was said to be very small (Q8). The percentage of time in which the laboratory's minicomputers are running was given in terms of 8-hour shifts per week (Q10):

DEC 11/70	11.5
DEC 11/34	5.0
Interdata	4.0

No answers were provided for QA1-QA4 and QB1-QB4. The higher-level language used is Fortran (QC1). The laboratory uses the SACC 360s under OS/MVT, the DEC 11/70 under RSX-11D, the DEC 11/34 under RSX-11M, and the Interdata under OS/GMT2 (QC2). As many as five tape drives per program may be required (QC4). Very often interactive processing is required (QC5). The number of library functions used was unknown (QC6). Several generalized service packages are used, such as scientific subroutines, statistical packages, matrix inversion, coordinate transformation, EBCDIC to ASCII conversion, Calcomp plot, and "other" plot packages (QC7). The total lines of code per program was unknown, but pre-existing code was estimated to approximate 20 percent (QC8). Programs generally consist of 15-20 percent of assembly language code (QC9).

7.5 Questionnaire - CR Composition

The CR composition system performs data reduction and analysis on data colcollected on IMP-H, I and J. No answers were provided for Q1 and Q2; in fact, Q2 appears to be irrelevant since minicomputers are not used in this area (Q10). All processing is done on the SACC 360s (Q3 and Q9).

It is the programmer's appraisal that the programs of the system are very poorly structured and that they would be of little use outside the CR Composition group (Q4). We assume that the programs evolved from IMP 6 to IMP 8, which would reflect reusability but not necessarily efficiency. At the same time, the programmer recognizes a potential for routines of general interest, given a restructuring of the programs and some standardization of the data. The candidates for commonality would be matrix generation and histogram programs, which contain standard techniques for cosmic ray telescope particle discrimination (Q5). Currently 98 percent of the software is unique to the IMP CR experiments (Q8).

The programmer appears to argue that 90 percent of the 3 x 10^{11} bits of data retained is sensor data (Q6) and 10 percent of the remainder is apparently used by other groups (Q7). The programmer calculates that software storage accounts for .012 percent of the data storage requirement (Q8).

The following bit-breakdown is offered as typical for the CR IMP data bases (QA1):

Experimental Data	90%
Data Flags	1%
Time	1%
Command	0%
Housekeeping	2%
Attitude	1%
Orbital	5%

Tape is the permanent storage medium for the data base (QA2) and 1 tape backup is taken of each tape (QA3). The programmer says that the tapes are kept indefinitely (QA4).

It appears that all data on the CR data base are used (QB1 and QB2), unlike the case of X-ray astronomy where a portion of the raw data was said to be extraneous. Nothing remarkable was performed with the data in the CR programs. It was said that "summary" programs reduce data by 10^2 (QB3).

The CR programs are written primarily in Fortran (QC1), with approximately 5 percent of a typical program in assembly language (QC9). Programs run at the SACC facility under OS/MVT (QC2). Programs use 150k - 300k of core storage; data presumably take an additional 50k of core storage. Approximately 10 tracks of temporary disk storage may be used (QC3), and 1 to 3 tape drives are used (QC4). No programs run interactively (QC5), and none are compiled during production runs (QC11).

It is said that no library functions are used (QC6); however, there may exist some confusion between "library functions" and "generalized service packages". Of the latter, it is said that "generalized I/O routines" are used, presumably FTIO. The CR programs also use packages for "bit manipulation" and "generalized plotting" (QC7). The system consists of approximately 30 programs, ranging in size from 800 to 3,000 lines of code (QC8). The programs evidently evolved, but no estimate of the amount of preexisting code was offered. Assembly language code is used for routines which "generate data bases, perform statistical analysis", compute averages, generate plots, and tabulate (QC10).

8.0 Conclusions and Recommendations

While a firm quantitative estimate of the dollar benefits and disadvantages of moving to a more centralized data processing for NASA's Space Science investigations cannot be reached on the basis of this general study, some definite conclusions can be reached. Centralizing the data base has some clear cost advantages. Such a data base precludes the need for making at least one copy, and possibly two copies, of all data, giving a cost savings for this process of $1.8 to 3.6 million. If the data return from Spacelab missions is 4 to 8 times 10^{12} bits, this savings could range from $2.4 million to $10 million. These advantages may be offset by the costs of data communications, amounting to up to $600,000 for 10^{12} bits transmitted. However, centralizing the data processing and colocating this function with the data base will eliminate the need for such massive data transmission. The net savings is hard to estimate at this time.

The other advantage of centralizing the data processing will be the tendency to encourage the reuse of coding among groups. Estimating an average of $10,000 per group for 100 investigator groups results in a possibility for saving up to $1 million in programming costs.

Since a quantitative, supportable estimate of cost savings resulting from a centralized SSDS including both the data base and the processing capability is practically impossible at this time, a prototype system would be an ideal method of demonstrating what these savings could be. With the relatively low hardware costs, such a prototype consisting of a superminicomputer could support approximately 25 users for an implementation cost on the order of $1 million. (See cost breakdown in NSSDC/WDC-A-R&S 79-02.) The experience gained with such a system, besides providing data on which to evaluate an overall SSDS, would provide valuable capabilities in its own right. Besides cost considerations involving data duplication, transmission, and storage, this experience would give information on the advantages and disadvantages of the centralized approach to the overall problem of information extraction from space science data. Many questions that are of necessity left unanswered in this report would be answered through experience.

This prototype would serve another function - it would be the first step in the proposed phase-in plan if the full system were to be implemented. Because interactive processing is new to many users, many system design parameters can be determined during this period. These include requirements for RAM storage, core, communications, graphics terminals, and software.

But most important, the prototype would establish user acceptance for the SSDS approach. Similar capabilities have been used with Coordinated Data Analysis Workshops with great success. This acceptance by the user community is absolutely essential, and the best way to achieve this is through the establishment of a track record showing that an SSDS can meet the needs of OSS investigators.

BIBLIOGRAPHY

ABR Abrams, Phyllis, "Facsimile: A Modern Business Tool," Office Product News, March, 1979.

AND Anderson, Howard, "Computer-Based Message Systems," The Office, November, 1978.

CUN Cunningham, Peter A., and Ely S. Lurin, "The Future of Value Added Network Services," Telecommunications, July, 1978.

DOR Dorn, Philip H., "1979 DP Budget Survey," Datamation, January, 1979.

DUN Dunn, Donald A., "Limitations in the Growth of Computer-Communication Services," Telecommunications Policy, June, 1978.

GAM Gamble, Robert B., "VAN Services in the U.S.," Telecommunications, July, 1978.

HIR Hirsch, Phil, "XTEN from Xerox," Datamation, December, 1978.

HOW Howell, Dave, "Optical Communications System," Electronic Products Magazine, September, 1978.

KAN Kane, John, "Fiber Optic Cables Computer with MW Relays and Coax," Microwave Journal, January, 1979.

KEM Kemp, Graham S., "VANs in Europe," Telecommunications, July, 1978.

KIM Kimbleton, S.R., and G. M. Schneider, "Computer Communications Networks: Approaches, Objectives, and Performance Considerations," ACM Computing Surveys, September, 1975.

KIN Kinney, Harrison, "Stich in Time," Think, November/December, 1977.

LUN Lundell, E. Drake Jr., "IBM Moves in Disks, Too," Computerworld, February 5, 1979.

LUR Lurin, Ely S., and Edward T. Metz, "Get Ready for VANs," Datamation, July, 1978.

LUS Lusa, John M., "Xerox Files FCC Petition for Nationwide Digital Network," Infosystems, January, 1979.

MAN Mandell, Mel, "Distributed Data Processing," Computer Decisions, July, 1978.

MAR1 Martin, James, Telecommunications and the Computer, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1969.

MAR2 Martin, James, Systems Analysis for Data Transmission, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1972.

MES Messier, Claire V., "A Computerized Electronic Mail System," The Office, November, 1978.

MON Moncrief, Frank J., "Lasers Gain on Microwaves for Solar Power Satellites," Microwaves, November, 1978.

PEE Peeples, Donald E., "Measure for Productivity," Datamation, May, 1978.

RIN Rinder, Robert, "ACS is Coming," Datamation, December, 1978.

SAN Sanders, Ray W., "Comparing Networking Technologies," Datamation, July, 1978.

SHE Scherr, A. L., "Distributed Data Processing," IBM Systems Journal, Vol. 17, No. 4, 1978.

SCH1 Schultz, Brad, "The Year of the Supermini ... A Look at What's Available," Computerworld, February 12, 1979.

SCH2 Schultz, Brad, "PE 32-Bit Supermini Tagged at Only $33,500," Computerworld, February 19, 1979.

SEA1 Seamon, John, "Implementing DDP Successfully - Part I," Computer Decisions, January, 1979.

SEA2 Seamon, John, "Implementing DDP Successfully - Part II," Computer Decisions, February, 1979.

SHA Sharpe, William F., The Economics of Computers, Columbia University Press, New York, 1969, 1970.

STA1 Staff Writer, "The Distributed Trend: Flexibility at Lower Cost," Modern Office Procedures, January, 1979.

STA2 Staff Writer, "United Nations Conference via Satellite and Facsimile," The Office, November, 1978.

STA3 Staff Writer, "General Dynamics Facsimile Network," The Office, November, 1978.

STA4 Staff Writer, "New Options in Communications Methods," Modern Office Procedures, January, 1979.

STA5 Staff Writer, "Cost for DP People Expected to Drop," Computerworld, February 19, 1979.

STA6 Staff Writer, "Optical Communications: Seasoned Systems Complement and Compete with Microwave Methods," Microwaves, November, 1978.

STA7 Staff Writer, Auerbach Reporter: Computer Technology Edition, Philadelphia, February, 1978.

STA8 Staff Writer, All About Plug-Compatible Disk Drives, Datapro Research Corp., Delran, New Jersey, October, 1977.

WAN Wanner, James F., "Wideband Communication System Improves Response Time," Computer Design, December, 1978.

WEL Wells, Larry J., "Facsimile: What It Is and How It Is Used," The Office, November, 1978.

WHI White, Wade, and Morris Holmes, "The Future of Commercial Satellite Telecommunications," Datamation, July, 1978.

YAS Yasaki, Edward K., "The High Cost to Maintain," Datamation, February 1979.

EXHIBIT A

PRECEDING PAGE BLANK NOT FILMED

Exhibit A - GSFC Science Software Questionaire

1. How much SACC computer time (in units of 360/91 hours; address modeling effort also) do you use per year per space flight instrument?

2. How much time (in the units of the minicomputers) do you use on computers (minis, etc.) in your own lab, per year per flight instrument?

3. Can we shift work to the SACC computers with no significant adverse effects?

4. What percent of the software developed in your laboratory in the same time period (data reduction, analysis, etc.) could probably be used by other groups? How many man-hours is this?

5. By how many other groups could the typical line of this common software be used?

6. Of all the data held on tape or other media by your group, what percent of it (and how many bits) results directly from your group's flight instruments (do not include time words, ephemeris words, etc.)?

7. By how many other groups is the typical bit of your non-sensor-output stored data held?

8. What percent of the software lines you store are unique to your group? What storage volume is this relative to your data storage?

9. What percent of the software you store is run only on SACC computers, on your own minis; on both?

10. What percent of the time arc your minis running?

Characteristics of Data Structures

A1. What is the fraction of bits in each category?

A2. What is the storage mechanism used (tape, disk pack, cassette, other)?

A3. How many back-up copies are made?

A4. How long is the data kept in this logical format?

Functions of a Program

B1. What does the program do to each type data in the input data structure?

B2. What other parts of the data structure does the program use in operating on a given type? (The program may only alter the experiment data but may use the data flags, housekeeping, and command to do this.)

B3. If a new form of machine-sensible data structure is an output, then answer questions A1-A4 and give the reduction or expansion factor (for bit storage) for each type of data.

B4. What is the amount and form of the other outputs? Briefly describe their purpose and characteristics.

Characteristics of a Program

C1. What higher level language is used?

C2. On what machines does the program run and what operating system is used?

C3. What are the core and disk requirements for a) program? b) data?

C4. What are the requirements for tape drives and output devices?

C5. Does the program require interactive input from the user?

C6. How many library functions are used?

C7. How many generalized service packages are used? What are they (e.g., utility - EBCIDIC to ASCII, plot - Calcomp, scientific package - matrix inversion)?

C8. How many total lines of code is the program? How much is preexisting specialized code (C3. above)?

C9. How much of the program is in assembly language?

C10. What does the code specifically developed for this do: curve fitting, least squares, numerical differentiation/integration, time series, analysis, distribution functions)?

C11. What portion of the program is normally compiled in a typical run?

Letter Establishing SSDS Study Group

December 12, 1978

TO: Distribution

FROM: 600/Director of Sciences

SUBJECT: Space Science Data Service Study Group

The following persons are constituted as a special study group to consider the possibility of a Space Science Data Service (SSDS):

Joseph King, Chairman
James Vette
Fred Shaffer
Edward Sullivan
William Mish
Gerald Muckel
Daniel Klinglesmith
Paul Smith
Peter Bracken
Charles Wende
Leo Davis
Michael Mahoney

The objective of a SSDS shall be to provide rapid and selective access to the data and to the computational capabilities necessary to carry out scientific investigations supported by the Office of Space Science to principal investigators, guest investigators, theorists, and other data users.

The SSDS is conceived presently in terms of one or more data facilities equipped and configured to support on a continuing basis all OSS-supported data analysis and knowledge extraction activities. The study group should determine one or more possible configurations of a space science data facility to meet the objective of the OSS Data Service, determine a realistic phase-in plan for its operation, and present data concerning initial installation and annual operating costs.

The Chairman of the study group will report progress in the study to me monthly at the Sciences Directorate Experiment Review. He will also report to the Center Director at the Monthly Program Study Report Review if requested to do so. Time spent by members of the study group on the study should be charged to UPN XXX-385-36-10-73.

G. F. Pieper

Tables and Map

Table 1 (Updated 12/79)

OSS GSTDN/TDRSS MISSIONS DATA RETURN ($\times 10^{12}$ BITS)

Mission	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988
AE 5	.075	.075	.057							
DE-A			.103	.251	.251	.188				
DE-B			.062	.150	.113					
CCE					.015	.057	.019			
COBE							.006	.025		
EUVE							.022	.072		
GRO							.261	.628	.471	
HEAO 2	.200	.200	.200	.085						
HEAO 3	.059	.200	.091							
IMP 8	.031	.025								
IRAS			.013	.028						
ISEE 1	.204	.204	.204	.204	.204	.154				
ISEE 2	.103	.103	.103	.103	.103	.079				
ISEE 3	.062	.062	.062	.062	.062	.062	.062	.062	.062	.062
IUE	.012	.012	.012	.012	.012	.012	.009			
OAO	.015									
OPEN EML								.437	.320	.200
OPEN GTL							.200	.200	.200	.200
OPEN IPL							.085	.085	.085	.085
OPEN PPL							.515	.801	.801	.801
SAS 3	.015									
SL 1				.120						
SL 2				.286						
SL 3				.286						
SL's					.858	.858	.858	.858	.858	.858
SM D/L			.057	.141						
SM D/M					.157	.267	.267	.200		
SME			.009	.031	.022					
SMM		.518	.565	.565	.565	.565	.047			
ST					.060	.729	.729	.729	.729	.729
UK 5	0.062	.047								
TOTAL	0.838	1.446	1.538	2.324	2.422	2.971	3.080	4.097	3.526	2.935

Table 2

OSS DSN MISSIONS

DATA RETURN (BITS)

Mission	1979	1980	1981	1982	1983	1984	1985	1986	1987
Pioneer 10	2x10^{9}	2x10^{9}	.2x10^{9}	.2x10^{9}	.2x10^{9}	.2x10^{9}	.2x10^{9}	.2x10^{9}	.2x10^{9}
Pioneer 11	3x10^{9}	.5x10^{9}	.2x10^{9}	.2x10^{9}	.2x10^{9}	.2x10^{9}	.2x10^{9}	.2x10^{9}	.2x10^{9}
Viking 1/2	32x10^{9}	---	---	---	---	---	---	---	---
Voyager 1	10^{12}	10^{12}	65x10^{9}	65x10^{9}	65x10^{9}	65x10^{9}	65x10^{9}		
Voyager 2	10^{12}	65x10^{9}	10^{12}	65x10^{9}	65x10^{9}	65x10^{9}	65x10^{9}	.9x10^{12}	
Pioneer Venus									
Orbiter	22x10^{9}								
Probe	5x10^{6}								
Galileo				.17x10^{12}	.23x10^{12}	.23x10^{12}	.89x10^{12}	1.1x10^{12}	.2x10^{12}
ISPM									
NASA					50x10^{9}	53x10^{9}	53x10^{9}	53x10^{9}	39x10^{9}
ESA					25x10^{9}	26x10^{9}	26x10^{9}	26x10^{9}	20x10^{9}
VOIR							118x10^{12}		
Helios 1/2	63x10^{9}	63x10^{9}	63x10^{9}	63x10^{9}					
Total (x10^{12}Bits)	2	1	1	0.3	.4	0.4	1.*	2	0.3
Non Imaging (x10^{12} Bits)	0.2	0.2	0.2	0.3	0.4	0.4	1.*	1	0.2

*Does not include VOIR

Table 3 (Updated 12/79)

CURRENT AND PROJECTED OSS DATA VOLUME PER CALENDAR YEAR
($x10^{12}$ BITS)

	1979	1980	1981	1982	1983	1984	1985*	1986	1987	1988
DSN NON-IMAGING	0.2	0.2	0.2	0.3	0.4	0.4	1.0	1.0	0.2	0.2
DSN IMAGING	1.8	0.8	0.8	0	0	0	0	1.0	0.1	0
DSN TOTAL	2.0	1.0	1.0	0.3	0.4	0.4	1.0	2.0	0.3	0.2
GSTDN/TDRSS	0.8	1.5	1.5	2.3	2.4	3.0	3.1	4.1	3.5	2.8
TOTAL	2.8	2.5	2.5	2.6	2.8	3.4	4.1	6.1	3.8	3.0

*DOES NOT INCLUDE VOIR

Table 4

EXPERIMENT NUMBERS BY CATEGORY

	Total No.	No. of Different Groups
Magnetometers	19	4
D. C. Electric Field	6	2
VLF, Plasma Waves	20	3
Plasma	38	12
Charged Particles	8	3
Cosmic Rays & H.E. Particles	38	15
Mass Spectrometers	11	3
Temp/Pressure/Density	15	5
Langmuir Probes/RPA	13	4
Radio Science, Tracking, Radar	32	6
V/UV/EUV	47	23
IR, Microwave, Radio Astronomy	21	9
X & γ rays	28	11
Micrometeorites	5	3
TV Imagery	9	5
Other Planetary Surface	5	3
Other Planetary Atmos.	9	4
Other	3	3
TOTAL	327	118

Table 5

OSS PI'S LISTED IN RAPSE

Affiliation	STDN PI's	DSN PI's	Total PI's	Affiliation	STDN PI's	DNS PI's	Total PI's
GSFC	44	20	64	MSFC	3	0	3
Iowa	12	11	23	Kitt Peak	0	3	3
JPL	3	17	20	USC	0	3	3
ARC	0	19	19	Princeton	1	2	3
MIT	8	9	17	U NH	1	2	3
U COL	8	7	15	U MD	3	0	3
Cal Tech	5	4	9	GISS	0	2	2
LMSC	7	1	8	LRL	2	0	2
UT Dallas	7	1	8	Columbia	2	0	2
UCLA	4	4	8	JHU	2	0	2
NOAA/SEL	7	1	8	SUNY, Albany	1	1	2
UC-Berk.	8	0	8	Florida State	0	2	2
Ariz.	1	6	7	Utah U	0	2	2
Chicago	4	3	7	BTL	0	1	1
Stanford	3	4	7	Wash U St. L.	1	0	1
LARC	0	6	6	Aerospace	1	0	1
TRW	2	4	6	HAO	1	0	1
Harvard/SAO	5	0	5	UT Austin	1	0	1
LASL	4	1	5	MDAC	1	0	1
NRL	5	0	5	Drexel	0	1	1
AFGL	5	0	5	SRI Int.	0	1	1
UCSD	3	2	5	Utah St U	1	0	1
Wisc.	2	3	5	Bell Aerospace	1	0	1
Mich.	5	0	5	Boston U	1	0	1
APL	3	2	5				
USGS	0	4	4	TOTALS	178	149	327

Table 6

REPRESENTATIVE NUMBERS OF OSS INVESTIGATORS BY AFFILIATION

AND CLUSTER LOCATION (FROM RAPSE)

Cluster Location	Affiliation	No.	Cluster Total
Los Angeles	JPL	54	
	Cal Tech	15	
	UCLA	14	
	TRW	7	
	USC	3	
	MDAC	2	
	CA State-Fullerton	1	
	SAI-Pasadena	1	101
Washington DC/GSFC	GSFC	71	
	APL	7	
	U MD	5	
	JHU	4	
	Smithsonian	1	88
CA Bay Area	ARC	26	
	UC-Berkeley	16	
	LMSC	14	
	Stanford	13	
	SRI	3	
	UC-SF	2	74
Boston	MIT	28	
	Harvard/SAO	18	
	BC	2	
	BU	2	50

Table 6a

REPRESENTATIVE NUMBERS OF OSS INVESTIGATORS BY AFFILIATION

AND CLUSTER LOCATION (FROM RAPSE) CONCLUDED

Cluster Location	Affiliation	No.	Cluster Total
Boulder	U Colorado	15	
	NOAA	15	
	NCAR/HAO	15	45
Cluster of Excellence - 1	U Wisc.	10	
	U Chicago	9	
	U Iowa	9	
	U Minn.	7	35
Tucson	U Ariz.	22	
	Kitt Peak	7	29
Cluster of Excellence 2	UT-Dallas	9	
	UT-Austin	7	16
San Diego	UCSD	11	
	Scripps	2	13
New York City	GISS	6	
	Columbia	4	10
Houston	JSC	5	
	Rice	2	
	U Houston	2	
	UT-Galveston	1	10
Ann Arbor	U Mich.	9	
	Bendix	1	10
CLUSTER TOTALS			482

Table 7

.REPRESENTATIVE NUMBER OF OSS INVESTIGATORS

BY AFFILIATION THAT ARE NOT CLUSTERED

Affiliation	No.	Affiliation	No.
MSFC	15	Pomona College	1
LARC	13	UT San Antonio	1
Cornell U	6	FSU	1
Utah State U	5	Baylor	1
Princeton U	5	U Virginia	1
U Hawaii	4	Ariz. St U	1
U NH	4	Grambling U	1
U Florida	3	Sandia Corp	1
U Wash.	3	U Alaska	1
Santa Barbara Res. Ctr.	3	Ohio State U	1
U Tenn.	3	U Utah	1
Brown U	2	Yale U	1
Drexel U	2	U Illinois	1
SUNY, Albany	2	SUNY, Buffalo	1
Wash. U St. L.	2	BTL	1
Lowell OBS	2	Bell Aerospace	1
Martin Marietta Denver	2	U NM	1
U Penn.	2	BYU	1
U Mass.	1		
NON-CLUSTER TOTALS			97

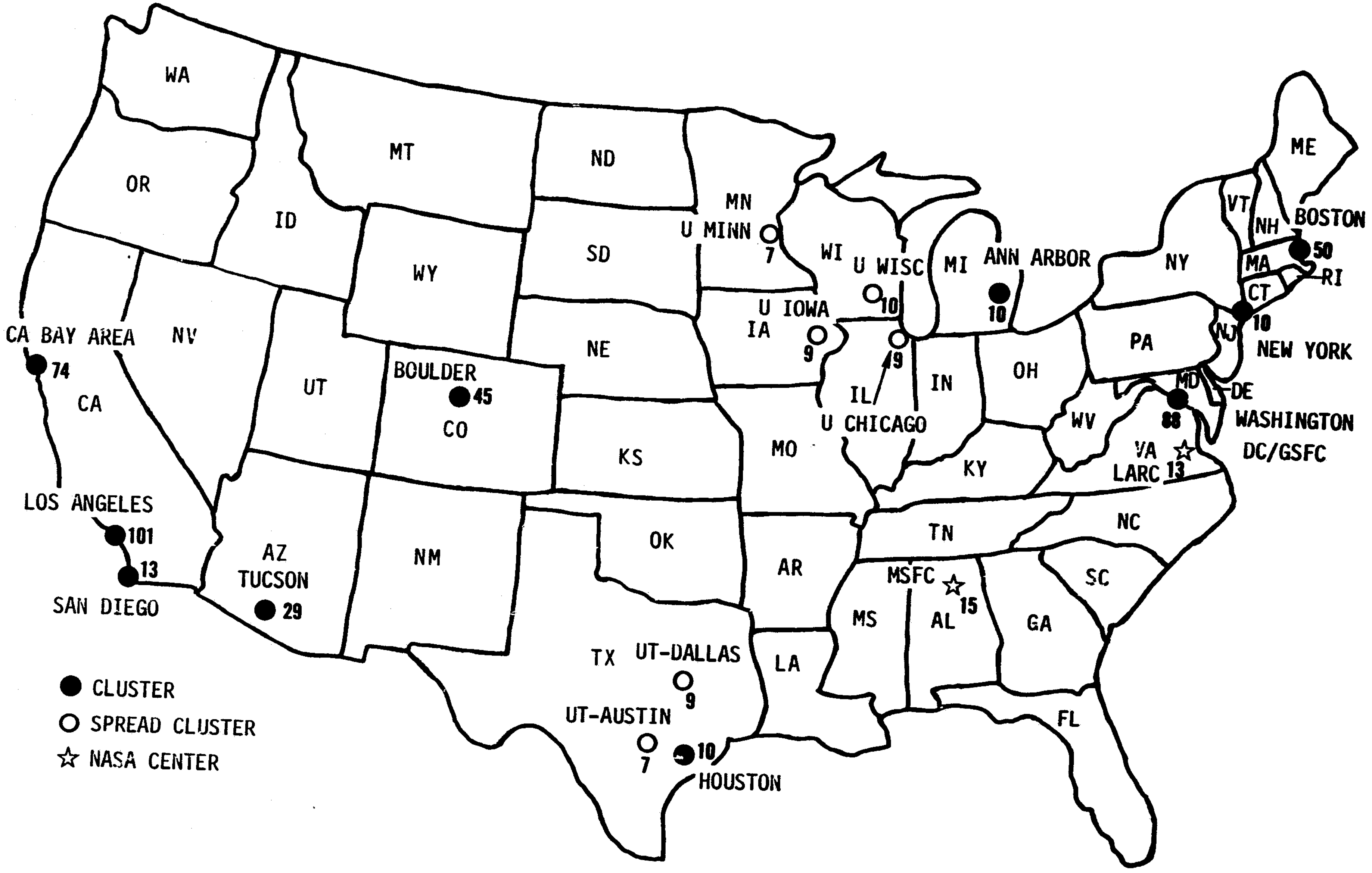

FIGURE 3. NUMBER OF OSS INVESTIGATORS BY CLUSTER